ISL 9028-37

TECHNICAL REPORT

# Automating Cataloging Functions in Conventional Libraries



15 APRIL 1963



INFORMATION SCIENCES LABORATORY

Itek

ITEK CORPORATION

LEXINGTON 73, MASSACHUSETTS

7

## ABSTRACT

Partial automation of processing functions has been achieved. A tape typewriter creates machine-interpretable bibliographic records, which are then processed on a specially designed digital processor, called the Crossfiler. This device makes use of a machine-interpretable natural format, which is a variable field, multilevel, sequential format, using natural text delimiters as machine-interpretable boundary codes. In printout, the Library of Congress format is simulated. The system is presently being implemented in a large Air Force library.

## FOREWORD

This technical paper on automating cataloging functions in conventional libraries was prepared by Paul J. Fasana of the Information Sciences Laboratory of Itek Corporation, Lexington 73, Massachusetts, and describes work performed under contract AF19(604)8438 for the US Air Force Cambridge Research Laboratories, Bedford, Massachusetts. The paper represents the efforts of the entire Itek project team that participated in the contract.

## INTRODUCTION

Two years ago, Itek Corporation was awarded a contract by the US Air Force to study automatic processing of large volumes of technical information. The prime objectives of the study were:

1. To examine the methods currently used by the Air Force Cambridge Research Laboratory (AFCRL) Research Library for the control of monographs, serials, and technical documents
2. To investigate the feasibility of a total system of mechanized processing routines
3. To achieve compatibility of the Research Library with machine systems in other information processing centers

Detailed studies and flow charts were made of acquisitions routines, cataloging activities, circulation methods, and technical report distribution.

The Research Library at Laurence G. Hanscom Field, Bedford, Massachusetts, is organized to serve the AFCRL scientific community. The laboratories perform basic research in the physical sciences, particularly in geophysics and electronic communication. The Research Library's collection comprises some 180,000 monographs and bound journal volumes, over 3,200 journal subscriptions, and 270,000 technical reports. Annual acquisition rate is 2,500 monographs, 30,000 issues of journals, and 25,000 technical reports.

This paper discusses design and development work done on the Research Library's monograph cataloging routines. It describes:

1. Machine-interpretable natural format (MINF) — an encoded bibliographic format which in printout is compatible with existing library records
2. Itek Crossfiler — a digital machine designed and built to automate cataloging routines
3. Crossfiler System — a number of processing routines centered around the Crossfiler and having as its ultimate objective a total systems integration of technical processing activities

## BACKGROUND

Many large libraries and information centers — and their number is increasing — are faced with problems in the production of bibliographic records. Every item acquired by a library must be cataloged and integrated into the library's existing collection. Cataloging involves the performance of several related routines by which to generate a number of bibliographic records. These records in turn are filed in various catalogs and authority files and are used to uniquely identify and control items in the library's collection. The most important single record generated in the cataloging process is the "unit catalog card."

The unit catalog card is a master record of the bibliographic information discovered about an item. Specific information is extracted from it to create sets of secondary bibliographic records called "catalog card sets."

Each 3 by 5-inch card in a catalog card set contains all the unit card information plus a unique tracing that is generated by extracting an item of information from the body of the unit card. These tracings are made to provide additional access points to a particular bibliographic item.

Libraries presently generate catalog card sets by:

1. Manually typing each card in the set
2. Typing tracings on duplicated unit cards

In both methods, original information contained on the unit card must be retyped, which in turn requires proofreading, correcting, etc., and the sole product of the input typing is a visible record which can be used only in a manual file.

After thorough study of the AFCRL Research Library's processing routines, it was decided that the production of catalog cards was a critical point in the work flow. Since the library is of a typical medium size, solutions of its problems are applicable to libraries in general. Therefore, a generalized approach to library automation was formulated, and the initial step in this approach was planned around the cataloging activity.

Systems engineers often fail to properly consider an important and, it would seem, obvious point in approaching the problem of automation in an established library or information center. Much effort has been expended by librarians in analyzing and organizing acquisitions, and in creating a card catalog to control and interpret this collection. The possibility of totally replacing either the human intelligence required to analyze materials or the card catalog itself with a computer or any other kind of machine is unlikely. The initial design for a mechanized library must be compatible with the existing system and must seek to supplement, rather than replace, traditional library routines. The automated system must allow for gradual changeover from manual to machine routines, until the most efficient degree of automation for a particular library has been achieved. The Crossfiler System attempts to satisfy these basic requirements.

In the Crossfiler System, catalog card and machinable record production are combined by using a tape typewriter. This method would allow the library to maintain the continuity and integrity of its card catalog and present work flow, yet take an initial step towards automation. From a single input typing, the Research Library now produces:

1. Catalog card sets to be used in its card catalog, and
2. A machine-interpretable record to be used in a machine file for a fully automated system

## MACHINE-INTERPRETABLE NATURAL FORMAT

Before any store of information can be automated, two basic machine requirements must be satisfied. Information must be machine-accessible and machine-interpretable. Making information machine-accessible is simple; data need only be translated from written or printed records into a machinable form (i.e., paper tape, magnetic tape, or punched cards). This can be accomplished by using such devices as key punches and tape typewriters.

Making information machine-interpretable is more complex; considerable preliminary planning and analysis are required. In the past, machine interpretability was achieved in one of two ways:

1. By fixed-symbol coding, in which each item of information is tagged with a special symbol; each symbol is used to signify one type of information (e.g., Δ = title-page information, ‡ = subject information)
2. By fixed-field coding, in which particular kinds of information are associated with a fixed position in the machine medium; each position is used to signify one kind of information (e.g., columns 20 through 30 = subject information)

Neither method is completely satisfactory. In the first method, a special symbol must be devised and used to identify every type of information to be encoded. The generation of these symbols becomes extremely cumbersome when many different kinds of information are to be distinguished. A library catalog card, for example, contains 30 or more distinguishable types of information (e.g., class number, book number, title, imprint). In addition, the symbols themselves are difficult to generate, since the number of printing codes that are available on standard key-input equipment is severely limited (47 printing codes for IBM equipment, and 88 printing codes for automatic typewriters). Fixed-symbol coding of data creates additional problems in printout. If the symbol codes are not printed out for each item of information, proofreading routines become cumbersome, requiring continual reference to some master record; if they are printed, the symbols constitute a cluttered record with reduced legibility.

In fixed-field coding, the amount and form of information included are critically affected by the constraints of the media being used. For example, the IBM card has 80 columns, with only one printing character allowed per column. Catalog cards contain from 300 to 500 characters; therefore, if the information is to be printed in full, a minimum of four punched cards must be allotted for every catalog card. This usually means that the information must be drastically abridged. But in either case, since a dedicated field is the basis of machine identification, that field must always be reserved on every card for its associated type of information, regardless of whether a particular item has that information. In catalog cards, for example, less that 40 percent of the books cataloged have a series statement. In fixed-field coding, this position would have to be left vacant for more than 60 percent of the cards.

In both methods of making information machine interpretable, processing of information for input is quite complex and places a great many restrictions on both the cataloger and the input keyer.

The machine-interpretable natural format (MINF) combines the best features of both the fixed-symbol and fixed-field methods, resulting in an encoding method especially suitable for library purposes. Natural typing manipulations are used as boundary codes, thereby avoiding the rigidity of fixed-field coding. Information is recorded in a fixed sequence, thereby avoiding the complexities of fixed-symbol coding. These features reduce input keying routines to the level of ordinary typing. Fig. 1 shows an example of a library catalog card translated into MINF. This bibliographic adaptation of MINF is designed to fulfill three basic requirements:

1. To simulate the appearance of the conventional catalog card
2. To allow interfiling to be compatible with existing card catalogs
3. To encode bibliographic data for machine manipulation and interpretation

As mentioned above, two devices are used to structure the format and identify the nature of the data: sequential position and boundary markers. In the following sections, a detailed description is given of the bibliographic format showing how these devices are used.

## SEQUENTIAL POSITION

Four paragraphs are used in the format; these are distinguished by combinations of carriage returns, tabulate shifts, and spaces. Paragraphs are typed in a fixed sequence. Each paragraph is defined so as to correspond with a block of bibliographic information on a Library of Congress card (cf Fig. 1).

1. Call number paragraph
   a. Class number statement
   b. Author number statement
   c. Main entry statement

2. Descriptive paragraph
   a. Title and title-page transcription
   b. Imprint statement

3. Collation and notes paragraph
   a. Collation statement
   b. Notes statement

**Brandt, Conrad.**

A documentary history of Chinese communism, by Conrad Brandt, Benjamin Schwartz and John K. Fairbank. Cambridge, Harvard University Press, 1952.

552 p. 24 cm. (Russian Research Center studies, 6)

1. Communism—China. I. Schwartz, Benjamin Isadore, 1916- (Series: Harvard University. Russian Research Center. Russian Research Center studies, 6)

A 52–9[illegible]0

Harvard Univ. Library
for Library of Congress [10]

---

951.05
B75

Brandt, Conrad.

A documentary history of Chinese communism; by Conrad Brandt, Benjamin Schwartz and John K. Fairbank. Cambridge, Harvard University Press, 1952.

552p. Russian Research Center studies [6].

Communism—China. Schwartz, Benjamin Isadore, 1916- Harvard University. Russian Research Center. Russian Research Center studies, 6.

Fig. 1 — Examples of Library of Congress catalog card layout and Crossfiler machine-interpretable natural format

4. Tracing paragraph
   a. Subject statement
   b. Added entry statement
   c. Series statement

Tracing and Nontracing Mode

Of the ten statements distinguished in the encoded format, only four are used for tracing purposes: 2a, 4a, 4b, and 4c. These statements are said to be in the tracing mode. When processed on the Crossfiler, a tracing will be automatically generated for each complete phrase included in each of these statements. All other statements are in a nontracing mode, so tracings are not pulled from them. Fig. 2 shows a catalog card set generated by the Crossfiler, with the tracings automatically pulled and reformatted.

BOUNDARY MARKERS

Cards, paragraphs, statements, and phrases are separated and distinguished by conventional typing operations, which serve both as normal punctuation in the printout and as machine-interpretable boundary codes in the punched tape record. The boundary combinations are reserved for these uses exclusively:

1. Cards are introduced or separated by a sequence of at least five carriage returns.
2. Paragraphs are separated by a sequence of one carriage return and two tabulate shifts.
3. Statements are separated by a sequence of three spaces.
4. Phrases are separated by a sequence of two spaces.

Fig. 3 shows the encoded format with the boundary markers used to define cards, paragraphs, statements, and phrases.

Abridging and Permuting Tracings

Two special characters are provided to abridge and permute any statement in a tracing mode. The semicolon is used to abridge any statement in the tracing mode, and frequently in the descriptive paragraph in order to set off incidental title-page information. Fig. 4 shows how the semicolon is used.

The permuting device is a nonprinting symbol which allows additional permuted tracings to be pulled for any statement in the tracing mode. This symbol is used most frequently in the title statement to create shortened tracings or catchword titles. Fig. 4 shows how the permuting bar is used to create additional tracings.

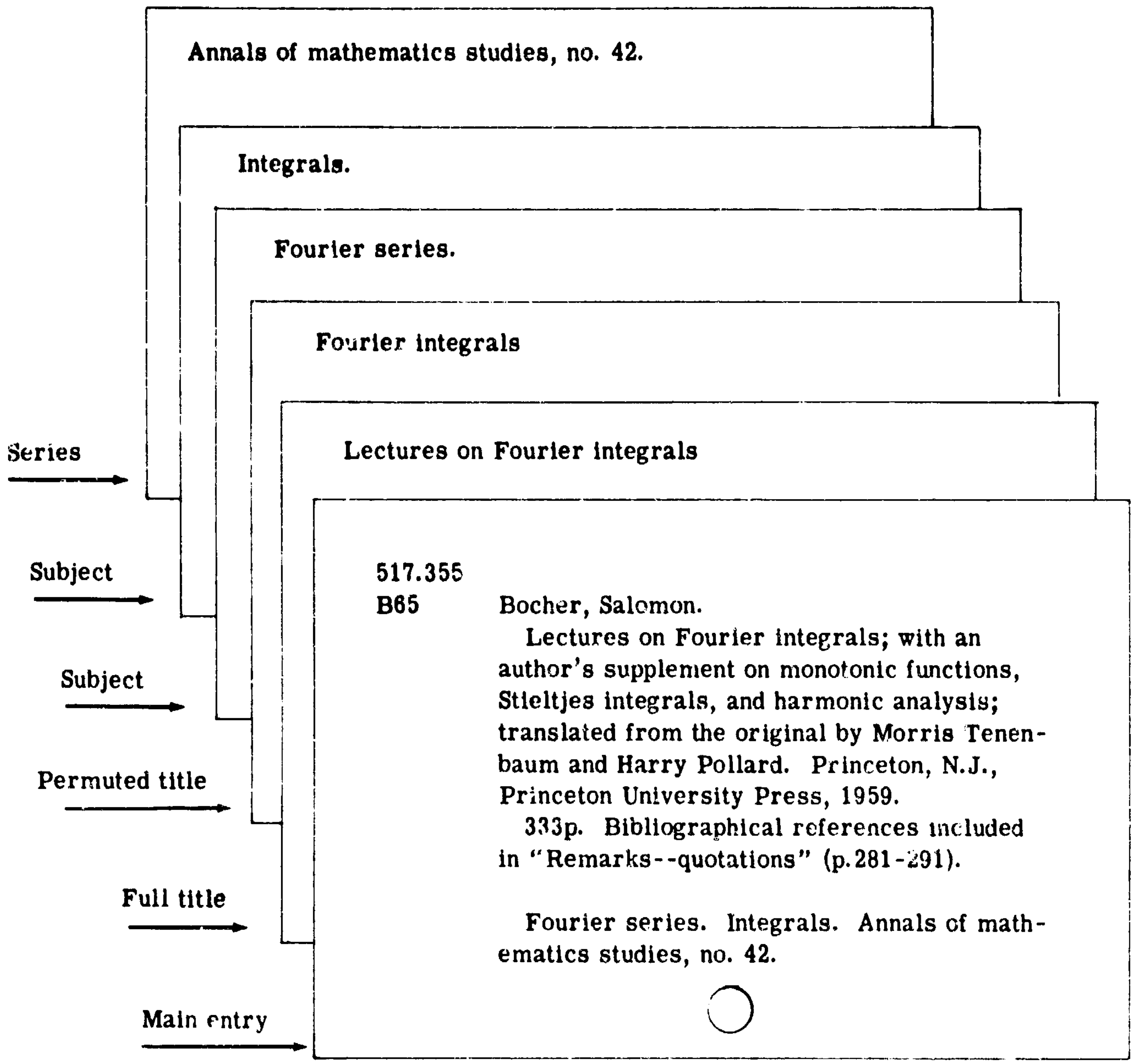


Fig. 2 — Crossfiler-generated catalog card set

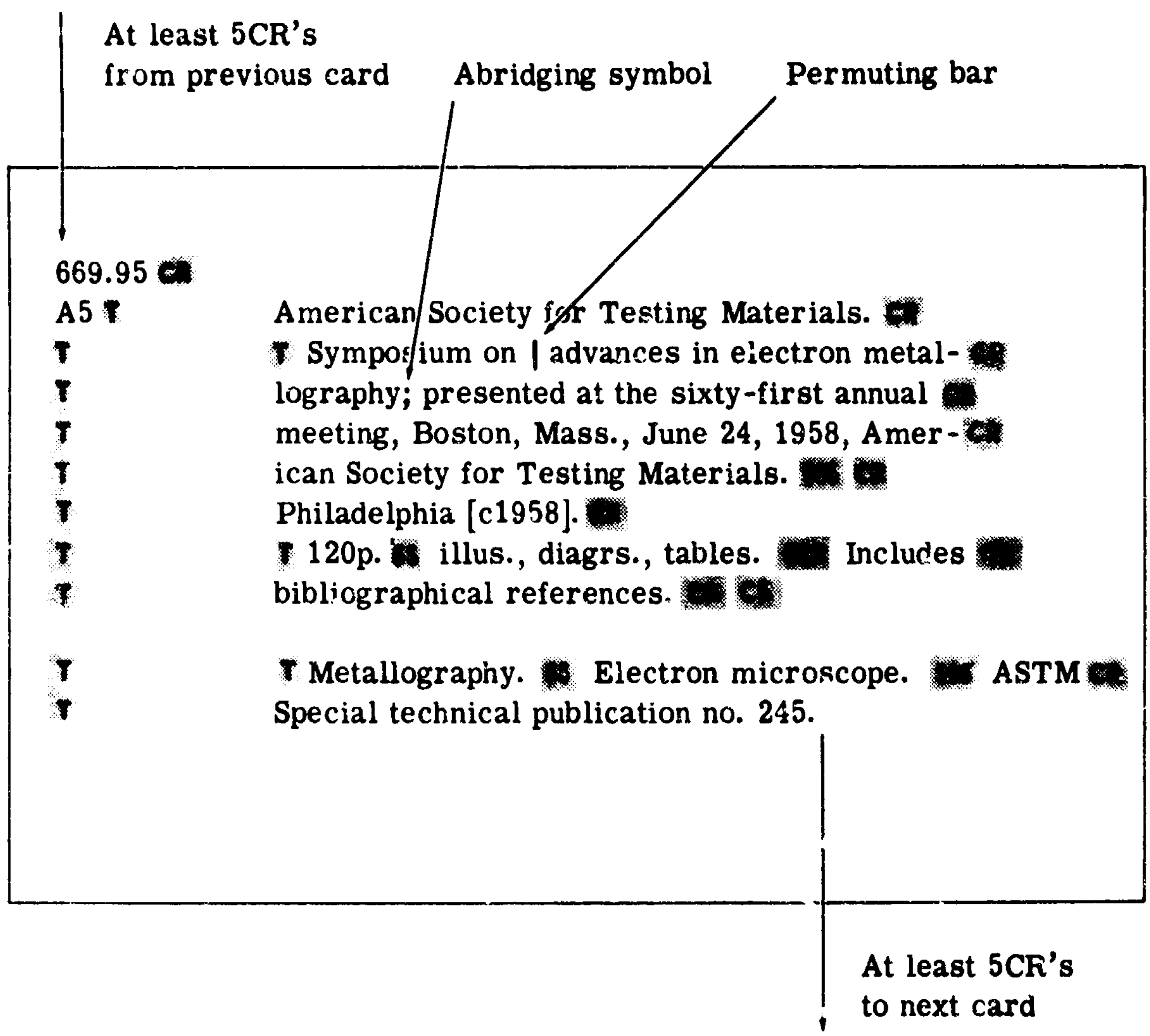


Fig. 3 — Crossfiler machine-interpretable natural format showing typing manipulations used as boundary codes

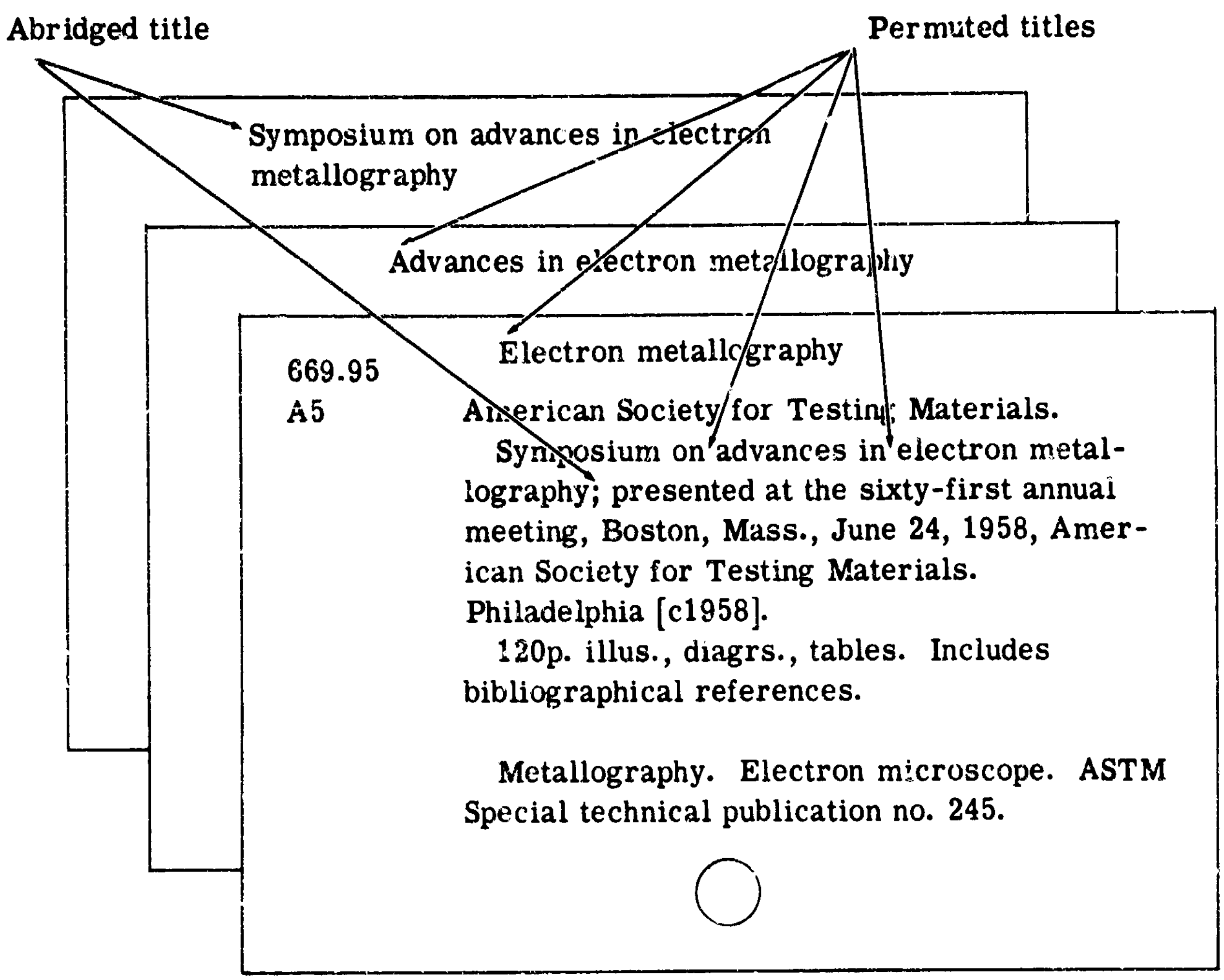


Fig. 4 – Abridged and permuted title entries automatically generated and reformatted by Crossfiler

## THE CROSSFILER

The Crossfiler* is a small-scale, transistorized digital processor designed and built especially for cataloging automation based on the MINF concept. It reads, interprets, and manipulates bibliographic data from a properly formatted, punched paper tape that represents catalog cards.

The primary function of the Crossfiler that is presently being installed at the AFCRL Research Library is to produce sets of diversely headed catalog cards. It accomplishes this by processing the punched paper tape representation of a unit catalog card and punching out a secondary tape representing a completed catalog card set. The expanded output tape is then loaded into an automatic typewriter and typed out on continuous-form card stock.

The unexpanded input tape can be reused to produce other library records, such as accessions lists, book-form indexes, book-pocket information, circulation records, etc. Since the tape format is designed to be compatible with general purpose digital computers, a mechanized retrieval file is automatically accumulated as a by-product of the catalog activity.

Operation of the Crossfiler is extremely simple; input tapes are loaded into the reader and start buttons are pushed. The machine then automatically processes the original tape document to produce an expanded output tape.

---

* Input tapes can be processed either on a general purpose computer or on the Crossfiler; this paper describes Crossfiler processing routines only.

## THE CROSSFILER SYSTEM

A number of input processing routines have been developed for use with the Crossfiler and its encoded format. These routines, together with Crossfiler operations, comprise the Crossfiler System (see Fig. 5). The input control device used in this system is a specially designed worksheet (see Fig. 6).

The following processing routines represent a practical application of the Crossfiler System and are presently being implemented in the AFCRL Research Library:

1. A worksheet is inserted in every book being processed before the book passes to cataloging. The cataloger transcribes bibliographic information into the appropriate blocks of the worksheet. Completing the worksheet amounts to format translating, since the worksheet is designed to sequence and position blocks of bibliographic data for the encoded format.

2. Completed worksheets are then passed on to the tape typist who transcribes the data from the worksheet. The necessary format boundary codes are inserted into the tape as part of the typing routine.

3. Initial typing is done on continuous form paper. Copy is read and corrected, the necessary corrections are made in the paper tape, and the tape is then ready to be processed on the Crossfiler.

4. Corrected tapes are expanded by the Crossfiler and these tapes are typed out on Flexowriters using continuous card stock with top and bottom cuts and side perforations. Original input tapes are stored for future use.

5. Catalog card sets are assembled and sent to be filed in the appropriate catalogs.

### CATALOGING COSTS

It is difficult at this time to give exact cataloging costs for the Crossfiler System, since it is still undergoing testing. However, statistics accumulated during the experimental phase indicate that significant savings in time and money will be realized once the system is fully operative. Two sets of figures are presented in the following paragraphs to substantiate this statement.

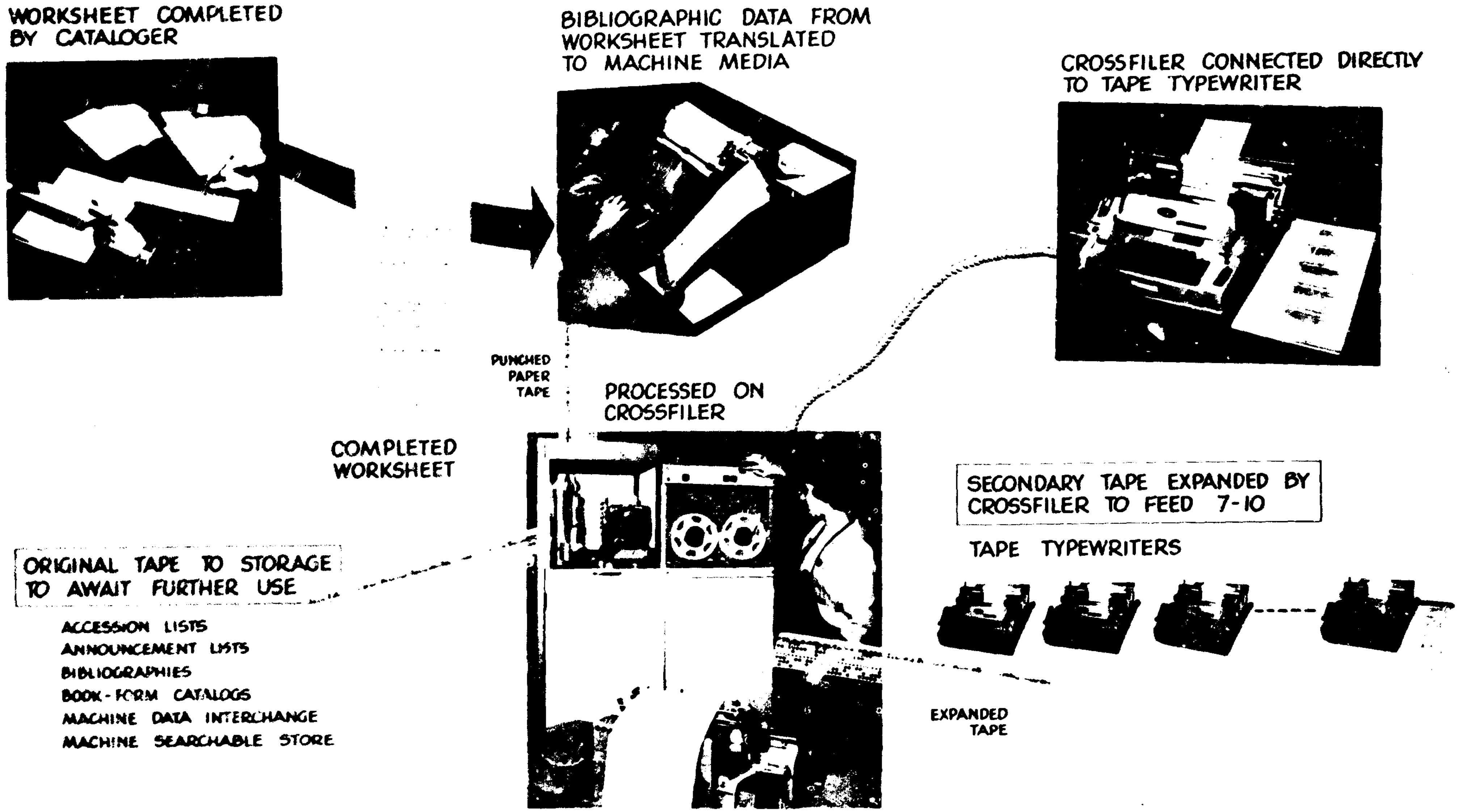


Fig. 5 — Crossfiler system: catalog card production

| BOUNDARY | DATA FIELD |
|---|---|
| CCCCC | Call Number 669.95 A5 |
| T | Main entry American Society for Testing Materials. null |
| CTT | Title and Title-Page Transcription Symposium on/advances in/electron metallography; presented at the sixty-first annual meeting, Boston, Mass., June 24, 1958. American Society for Testing Materials. |
| SSS | Imprint Philadelphia [c1958] |
| CTT | Collation 120p. illus., diagrs., tables. |
| SSS | Notes Includes bibliographical references. null |
| CTT | Subjects Metallography Electron microscope |
| SSS | Added Entries null |
| SSS | Series ASTM Special technical publication no. 245. null |

Fig. 6 — Worksheet for monograph cataloging

## Cataloging Time

A sample of 2,500 monographs was cataloged to test various aspects of the proposed system. Cataloging was done by a single cataloger and included classifying, descriptive cataloging, and subject analysis for each item. The test sample was divided into two general groups:

1. Monographs with Library of Congress information available in the form of LC cards, or photographic reproductions of National Union Catalog entries (60%)
2. Original cataloging with little or no supplementary information (40%)

The cataloging rate for the first group was 7 monographs per hour, for an average of 8.5 minutes per monograph. The cataloging rate for the second group was 4.5 monographs per hour, for an average of 15 minutes per monograph. Overall, the rate was 6.8 monographs per hour, for an average of 10 minutes per monograph. This figure, when compared with the AFCRL Research Library's present cataloging average of 0.5 monograph per hour, is significant.

Two factors account for the significantly higher cataloging rate achieved in the test sample. First, books were assembled into groups of one hundred and processed in a strict sequence. Each routine was designated either as a professional or a clerical task. For example, authority-list checking was considered essentially a clerical task and was performed by a clerk before the books were given to the cataloger. This grouping of books and sequencing of routines freed the cataloger from a great many clerical activities.

The second factor involved extensive use of book-form authority lists. A great deal of a cataloger's time is spent consulting card authority lists which take him away from his desk and interrupt his train of thought. By providing the cataloger with a book-form copy of the authority list, a considerable amount of his time was saved.

## Catalog Card Production Costs

The clerical cost of preparing and filing a catalog card in the AFCRL Research Library is 18.5 cents per card, or approximately $1.30 per card set.* This figure is based on the use of Library of Congress cards, and includes the cost of ordering, typing headings, proofing, sorting, and filing. It does not include cataloging costs. Using the Crossfiler System, the cost is 11 cents per card, or 77 cents per card set. This cost includes typing, materials, sorting, and filing. None of the figures quoted include machine costs or overhead.

---

* An estimate from a study in progress being conducted by Mr. Richard Snyder at Massachusetts Institute of Technology Library places the cost of producing a catalog card in a similar manner at 17 cents per card. This figure includes the cost of the LC card, 5.32 cents; typing headings, 4.22 cents; filing and sorting, 7 cents. It does not include the cost of ordering the card.

Another interesting comparison involves the amount of handling time required in both systems. Approximately 28 minutes of clerical time per card is presently needed to order, prepare, and file a card in the AFCRL Research Library. This figure does not include delay or storage time. In the Crossfiler System, handling time is cut to about 2.5 minutes per card. This time includes that for initial input typing, proofing, processing, typing out, and filing. This significant saving in handling time is a result of cards being immediately ready for filing once they have been typed out on the tape typewriter.

Machine Costs

The Crossfiler System is amenable to both computer and Crossfiler application. However, the Crossfiler was built for location within a library and for operation by unskilled personnel. It does a limited number of functions at a lower cost and with greater efficiency than a general purpose digital computer. The feasibility of producing commercial Crossfiler models is being studied, and preliminary estimates indicate that the price of production models will be in the twenty-to-thirty thousand dollar range.

At first glance, this figure seems to indicate that a Crossfiler would be economical only for extremely large libraries, but such is not necessarily the case. The speed of the Crossfiler allows it to process an item, i.e., the tape representation of a single document, and produce expanded output tape at the rate of one item per minute, or between 400 and 500 items per day. The high processing speed of the Crossfiler means that it can handle the total cataloging volume of several libraries. Since the machine accepts an input tape to produce a secondary expanded output tape that can be run on any compatible tape typewriter, several libraries can cooperate to use a single Crossfiler, thus sharing the capital and labor costs.

Such a cooperative venture would be simple to set up and efficient to run. Each library would prepare its own input tapes and send them to a central processing center; each would receive back an expanded output tape which the library would type out on its own tape typewriter.

## ADDITIONAL FEATURES

The by-product value of the machinable input record produced for the Crossfiler System has not been discussed in this paper. However, this machinable record should prove invaluable by producing a variety of bibliographic products on the Crossfiler without further human effort, including:

1. The production of announcement and accessions lists by automatically abridging and rearranging input records for various output printing formats
2. The production of control records for bindery routines and circulation activities
3. The production of intermediate data to be sorted and published as book-form catalogs with alphabetical indexes
4. The production of selective lists according to input category codes or class numbers

It should be stressed that the machinable input record is not confined to Crossfiler manipulation but is compatible with general purpose computers: Several of these applications are currently being investigated at Itek.

1. Generation of a machine-searchable file to produce demand bibliographies
2. Automatic compilation of printed subject bibliographies with various indexes, i.e., author, title, subject, etc.
3. Automatic compilation and typesetting of book-form catalogs with various indexes